J. Mol. Biol. (1994) **235**, 1532–1547

Three-dimensional Solution Structure of the Thrombin-binding DNA Aptamer d(GGTTGGTGTGGTTGG)

Peter Schultze, Román F. Macaya†, and Juli Feigon‡

Department of Chemistry and Biochemistry and Molecular Biology Institute University of California, Los Angeles, CA 90024, U.S.A.

The DNA oligonucleotide d(GGTTGGTGTGGTTGG) (thrombin aptamer) binds to thrombin and inhibits its enzymatic activity in the chain of reactions that lead to blood clotting. Two-dimensional ^{1}H NMR studies indicate that the oligonucleotide forms a folded structure in solution, composed of two guanine quartets connected by two T-T loops spanning the narrow grooves at one end and a T-G-T loop spanning a wide groove at the other end. We present the assignment strategy used, methods for the structure determination, and the refined three-dimensional structure of the thrombin aptamer. The initial structures were generated by metric matrix distance geometry using distance and dihedral bond angle constraints from NOE and coupling constants, respectively, and refined by restrained molecular dynamics and direct NOE refinement. Knowledge of the three-dimensional structure of this thrombin aptamer may be relevant for the design of improved thrombin-inhibiting anti-coagulants with similar structural motifs.

Keywords: two-dimensional NMR; DNA structure; quadruplex; G-quartet; G-DNA

1. Introduction

The sequence of the title molecule d(GGTTGGTGTGGTTGG) was originally identified by Bock *et al.* (1992) as the result of a selection project for DNA sequences that bind to thrombin and inhibit its enzymatic activity in the chain of reactions that lead to blood clotting. It constitutes the example with the highest binding constant out of a set of similar oligonucleotides with the consensus sequence d(GGtTGGN$_{2\text{-}5}$GGtTGG), where capital G and T are invariant nucleotides, lower case letters are variable but with a bias for a particular nucleotide, and N is any nucleotide. The procedure used to enrich for and isolate oligonucleotides with elevated ligand-binding capabilities from a pool of synthetic DNA or RNA with random sequences is called SELEX (Irvine *et al.*, 1991; Tuerk & Gold, 1990) or *in vitro* selection (Szostak, 1992) and the products of these selections have been termed aptamers (Ellington & Szostak, 1990). In this example, the fraction of thrombin-binding oligonucleotides was enriched by repeated passes through an affinity column with immobilized thrombin, each time followed by a polymerase chain reaction step on the isolated pool of bound sequences, and finally by cloning and sequencing of the enriched pool of oligonucleotides with high binding affinity for thrombin (Bock *et al.*, 1992).

The thrombin aptamer is the first example of a potential nucleic acid therapeutic agent targeted to a protein that does not physiologically bind nucleic acids. The binding site on thrombin was recently localized to the anion binding exosite (Wu *et al.*, 1992). Although the binding constant is only around 0·2 μM (Bock *et al.*, 1992; J. Feigon, unpublished results), it represents a promising lead compound as an anticoagulant.

We (Macaya *et al.*, 1993) and Wang *et al.* (1993) recently reported NMR studies that indicated that the thrombin-binding aptamer d(GGTTGGTGTGGTTGG) forms a unimolecular quadruplex in solution consisting of two G-quartets (Guschlbauer *et al.*, 1990; Smith & Feigon, 1992) connected by two T-T loops spanning the narrow grooves at one end and a T-G-T loop spanning a wide groove at the other end (*Fig. 1*). *T4* and *T13* are stacked over the neighboring G-quartet, positioned so that they can form a base-pair. Thus, all of the invariant bases in the consensus sequence for the thrombin aptamer are determinants of the tertiary structure. Here we present the assignment strategy used, methods for

† Present address: PharmaGenics, Inc., 4 Pearl Court, Allendale, NJ 07401, U.S.A.

‡ Author for correspondence.

0022-2836/94/051532-16 $08.00/0

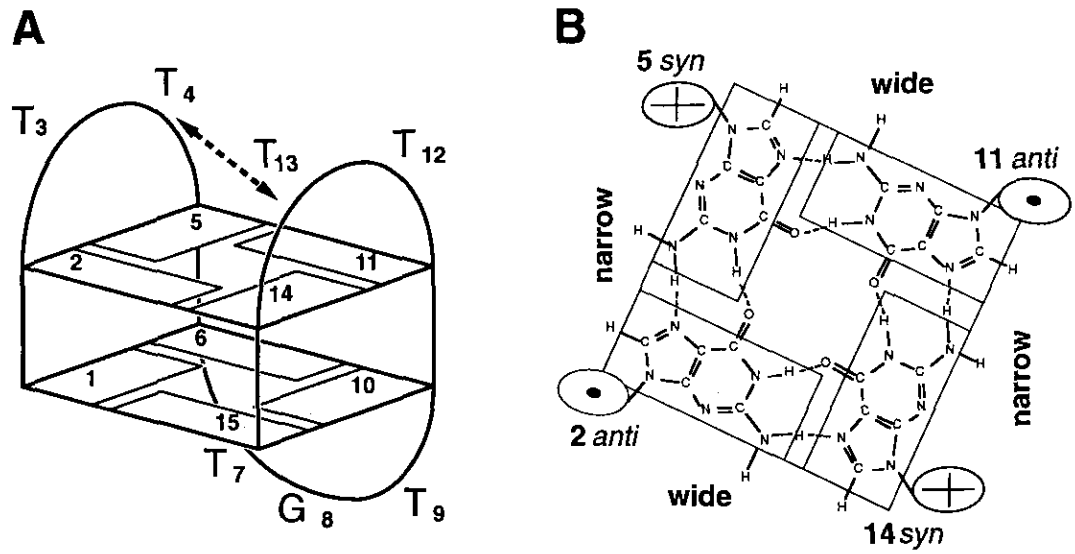

Figure 1. A, Schematic model of the structure of the thrombin aptamer. *syn* nucleotides are shaded. Model is the same as that of Wang *et al.* (1993) except for the interloop T·T base-pair, indicated by arrow at top. (B) Base-pairing structure of a guanine quartet. Numbering is the same as the top quartet in the model structure. Locations of the 2 wide and 2 narrow grooves are indicated.

the structure determination, and the refined three-dimensional structure of the thrombin aptamer. The initial structures were generated by metric matrix distance geometry using distance and dihedral bond angle constraints from NOE and coupling constants, respectively, and refined by restrained molecular dynamics and direct NOE refinement. The three-dimensional structure of the thrombin aptamer can be used as a starting point for generation of improved thrombin-inhibiting anti-coagulants with similar structural motifs.

2. Methods

(a) *Sample preparation*

The DNA oligonucleotides were synthesized and purified, and NMR samples were prepared, as previously described (Feigon *et al.*, 1992; Smith & Feigon, 1992), except that samples were ethanol precipitated with 1 M KCl rather than 1 M NaCl. The DNA oligonucleotides used in this study were: d(G1G2T3T4G5G6T7G8T9-G10G11T12T13G14G15; (thrombin aptamer) and the inosine, adenine, deoxyuracil and methyl C derivatives d(GITTGGTGTGGTTGG); **I2**, d(GGTTGITGTGGTTGG); **I6**, d(GGTTGGTITGGTTGG); **I8**, d(GGTTGGTGTGITTGG); **111**, d(GGATGGTGTGGTTGG); **A3**, d(GGTUGGTGTGGTTGG); **U4**, d(GGT**m^{5}C**GGTGTGGT**m^{5}C**GG); **C4C13** and d(GGTUGGTGTGGT**m^{5}C**GG); **U4C13**. NMR samples of the thrombin aptamer were 2·3 mM strand, 110 mM KCl, pH 6·1 (adjusted with KOH) in 400 μl ^2H_2O or 90%H_2O/10%^2H_2O, unless otherwise indicated.

(b) *NMR spectroscopy*

^{1}H NMR spectra were obtained at 500 MHz on a General Electric GN500 spectrometer. One-dimensional spectra in H_2O were acquired using a 1$\bar{1}$ spin-echo pulse sequence (Sklenář & Bax, 1987). NOESY† spectra in H_2O were acquired in the hypercomplex phase-sensitive mode (States *et al.*, 1982) using the standard pulse sequence (Kumar *et al.*, 1980) with the last pulse replaced by a 1$\bar{1}$ spin echo pulse sequence (Sklenář & Bax, 1987). Phase-sensitive NOESY spectra in ^2H_2O were acquired with presaturation of the residual H_2O peak during the recycle delay. P.COSY spectra were acquired with a flip angle mixing pulse of 90° (Marion & Bax, 1988). HOHAHA experiments were run with a mixing time of approx. 100 ms using MLEV17 (Davis & Bax, 1985). ^{31}P-^{1}H heteroCOSY spectra were acquired as described by Sklenář *et al.* (1986). One-dimensional spectra were processed with the GE NMR software (GEM16). Two-dimensional spectra were processed on a Silicon Graphics 4D/25 Iris work station using the program FTNMR and FELIX (Hare Research). Acquisition and processing parameters are given in the figure legends.

(c) *Assignments*

Assignments of the deoxyribose spin systems were obtained from analysis of P.COSY (Marion & Bax, 1988), and HOHAHA (Davis & Bax, 1985) spectra using standard methods (Feigon *et al.*, 1992; Wüthrich, 1986). Non-exchangeable base proton resonances were identified in NOESY spectra (Kumar *et al.*, 1980). Sequential assignment of the nucleotides (Feigon *et al.*, 1992; Wüthrich, 1986) was complicated by breaks in the sequential connectivities after T4, T7, T9 and T13, due to

† Abbreviations used: NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; COSY, correlation spectroscopy; HOHAHA, homonuclear Hartmann Hahn spectroscopy; TMP, trimethylphosphate; P.COSY, purged COSY; P.E.COSY, primitive exclusive COSY; p.p.m., parts per million; **I2**, etc., inosine-substituted derivatives; RMSD, root-mean-square difference; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate.

the presence of *syn* G nucleotides and loops. Sequence-specific assignments of non-exchangeable resonances were therefore obtained in 2 ways: (1) from analysis of NOESY spectra of aptamer derivatives (Macaya *et al.*, 1993); and (2) by analysis of ^{31}P-^{1}H COSY, as discussed in Results along with the assignments of the exchangeable resonances. Chemical shifts are referenced to external DSS for ^{1}H and TMP for ^{31}P, at the same salt and temperature conditions as for the sample.

(d) *Distance constraints*

Integrated peak intensities from NOESY spectra of the sample in $^{2}H_2O$ collected with mixing times of 50, 100, 150 and 250 ms were used in the structure calculations. For the starting structures, 188 distance constraints were extracted from peak integrals in the 50 ms mixing time NOESY spectrum. The integrated intensity of the most intense H2′–H2″ crosspeak (I_o) was used for distance calibration and set to 1·9 Å (r_o), and the intensities of all other crosspeaks were converted to distances using the $I = I_o r_{ij}/r_o)^{-6}$ relation, where I = peak integral and r_{ij} = distance between i and j protons. In all cases the lower bound was set at 1·9 Å and the upper bound at the calculated distance plus 0·5 Å. Thirty-six additional distance constraints were obtained from crosspeaks in a NOESY spectrum of the sample in H_2O collected with a 160 ms mixing time. The crosspeak intensities from this spectrum were classified visually as strong, medium or weak and converted to distances with upper limits of 3·0 Å, 4·5 Å and 5·5 Å, respectively. The average number of distance constraints per residue was 16.

Sixteen hydrogen bond constraints (1·9 Å) for the 2 guanine quartets with ideal H-bonding geometry were included in the structure calculations. In addition, for 2 of the 3 sets of calculations, constraints were included for an ideal T·T base-pair involving in one case the O2 atoms and the other case the O4 atoms as acceptors. To achieve close to ideal geometry for a T·T base-pair with a [C=O···H] angle of 120°, the following distances were used: [H-3-O2/4] = 1·9 to 2·0 Å, [H3–H3] = 2·3 to 2·6 Å, and [O2/4–O2/4] = 3·5 to 3·8 Å. In the 3rd set of calculations, no explicit T·T hydrogen bond constraints were included.

(e) *Dihedral bond angle constraints*

Homonuclear coupling constants $J_{1'2'}$, $J_{1'2''}$, $J_{2',3'}$, $J_{2'',3'}$, were determined from a P.COSY spectrum by an automatic optimization of the correlation between experimental and simulated data points covering the relevant crosspeaks using the program CHEOPS (Macaya *et al.*, 1992; Schultze & Feigon, unpublished results). These coupling constants were used as input for the PSEUROT program (de Leeuw & Altona, 1983; van Wijk *et al.*, 1992) and fitted using a 2-state model with 2 pseudorotation phase angles, P_N and P_S, and the ratio of the 2 conformations (see Table 2, below). The pucker amplitudes were kept constant at 38°. In the structure calculations, the deoxyribose conformations were restrained using dihedral angle restraints for ν_1 and ν_2 obtained from the PSEUROT results for the major conformational state for each residue.

Coupling constants $J_{P\text{-}H5'}$, $J_{P\text{-}H5''}$ and $J_{P\text{-}H3'}$ were obtained by simulation of the crosspeaks in a ^{1}H-^{31}P heteroCOSY spectrum using CHEOPS. These were used to constrain the backbone angles β and ε, respectively. The dihedral bond angles were estimated using the Karplus relation as modified by Altona and coworkers (Lankhorst *et al.*, 1984). Restraints for γ were estimated as described by Kim *et al.*, 1992) using the sum of couplings to H4′ in conjunction with the absence of strong H6/8–H5′,H5″ NOE crosspeaks. The $\Sigma J_{H4'}$ were determined by measuring outer peak splitting of the H3′–H4′ crosspeaks in a P.E.COSY experiment.

syn and *anti* residues were distinguished on the basis of the intensity of the intranucleotide H8/6–H1′ crosspeaks. Loose restraints of $-140 \pm 40°$ and $70 \pm 40°$ for the χ angle of *anti* and *syn* residues, respectively, were applied, mainly to increase the yield of structures with the correct *syn*/*anti* alternation in steps (1) and (2) of the calculations.

Since no direct experimental evidence was obtained for α and ζ, these angles were left unrestrained.

(f) *Structure calculations*

All structure calculations were performed with X-PLOR version 3.1 (Brünger, 1992). The whole sequence of calculations involved 4 steps: (1) substructure embedding using the metric matrix distance geometry algorithm; (2) completion of the partial co-ordinate sets by template fitting and regularization of this initial set of embedded structures by simulated annealing; (3) refinement of the best converged structures by simulated annealing and minimization; and (4) relaxation matrix refinement (Nilges *et al.*, 1991). The refinement protocol followed essentially the procedures outlined in X-PLOR manual, version 3·1; essential details and changes are described in the following.

In step (1), a substructure embedding was performed for 20 independent structures using the following 8 atoms per base: P, C1′, C3′, C5′, C2, C4, N3 and N1/N9. A complete set of atomic co-ordinates was generated in step (2) by replacing the partial set of atoms with a standard conformation for each nucleotide using a best fit superposition of the known atoms. This template fitting procedure was performed for both the initial embedded co-ordinates and their mirror image. After 200 cycles of energy minimization and a short period of molecular dynamics at high temperature, the total energies were compared for each pair of structures and the enantiomer with the higher energy was discarded. In about 30% of the structures this procedure yielded the incorrect enantiomer with a left-handed tetrahelix. These incorrect enantiomers always had significantly higher energies after step (3) and were discarded at that point. In the early calculations it also became clear that the original force field parameters did not guarantee the correct chiral configuration on the deoxyribose carbon atoms. Therefore, an additional set of "dihedral angle" restraints involving all 4 ligands on each chiral center was introduced throughout all calculations, with a high penalty for wrong configurations. For example, for the correct configuration at C1′, the "dihedral angle" O4′–N1/9–C2′–H1′ was restrained to $70 \pm 20°$ with a high force constant. The final phase of step (2) consisted of a slow-cooling dynamics calculation with subsequent energy minimization.

The structures were further refined in step (3) with a simulated annealing run in which the starting temperature of 2000 K was lowered to 100 K in 20,000 cycles. The lowest energy structure was used in a grid search for the correlation time that gave the best agreement between calculated and observed NOE intensities. This gave an overall correlation time of 4.0 ns. This value was used in the subsequent relaxation matrix refinement (step (4)) of the 6 structures with the lowest overall energies in step (3). The relaxation matrix refinement protocol consisted of a simulated annealing run from 1000 K to 75 K in 1000 cycles followed by energy minimization.

(g) *Calculation of helical parameters*

For the determination of helical twist and rise, the center of mass for all base atoms in each of the 2 tetrads was calculated, setting all atomic masses to unity. This was used to define the helical axis. The helical rise was defined as the distance between these centers. The helical twist was defined from: (1) the positions of the C1′ atoms or (2) the geometric centers of the consecutive G residues. These definitions give 2 different values.

3. Results

(a) *Imino proton spectra of thrombin aptamer and derivatives*

One-dimensional spectra of the imino proton resonances of the thrombin aptamer and derivatives in 110 mM KCl are shown in Figure 2. The spectra of the thrombin-binding aptamer in 110 mM KCl (Fig. 2A) and in the salt conditions of the selection buffer (140 mM NaCl, 5 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$) are essentially identical (see Macaya *et al.*, 1993). There are eight G imino resonances corresponding to the eight hydrogen-bonded guanines in the two G-quartets, as well as a high-field shifted G imino proton resonance from a non-hydrogen-bonded G. The imino proton assignments of the thrombin aptamer are indicated in Figure 2A. All of the aptamer derivatives also have eight hydrogen-bonded imino resonances from guanines and inosine, and the NOESY spectra in H_2O and ^2H_2O have crosspeak patterns similar to those observed for the thrombin aptamer, indicating that they form quadruplex structures with the same topology. For each of the inosine-containing aptamer derivatives **I2**, **I6**, and **I11** (Fig. 2B, C and E), a hydrogen-bonded inosine imino resonance appears at low field, ~16 p.p.m. In the **I8** derivative (Fig. 2D), a low-field imino resonance (~11 p.p.m.) corresponding to G8 disappears; the non-hydrogen bonded I8 imino resonates at ~11·5 p.p.m. The **A3** derivative (Fig. 2F) was one of the aptamer sequences that was found in the selection for thrombin binding DNA ligands (Bock *et al.*, 1992); the nucleotide at this position in the consensus sequence is variable. All of the derivatives except **U4** and **I8** form structures with lower stability than the thrombin aptamer, as judged by NMR temperature studies (data not shown). The aptamer derivatives were used to confirm the sequence specific assignments of the thrombin aptamer and to establish the topology of the thrombin aptamer.

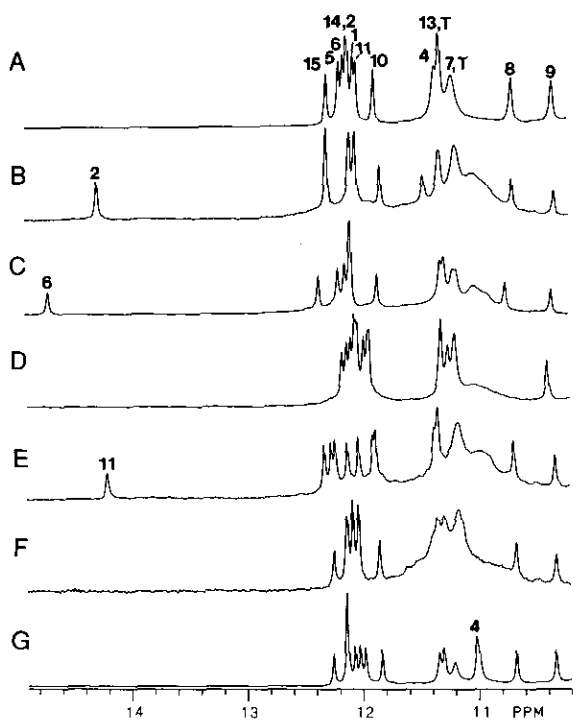

Figure 2. 500 MHz 1-dimensional ^1H NMR spectra of the imino resonances at 1 °C of: (A) thrombin aptamer; (B) **I2**, (C) **I6**, (D) **I8**, (E) **I11**, (F) **A3**, and (G) **U4** derivatives. Samples were 2·0 to 2·3 mM strand in 400 μl of 90%H_2O/10%^2H_2O, 100 mM KCl (pH 6·1). Spectra were acquired using a $1\bar{1}$ spin echo pulse sequence (Sklenář & Bax, 1987), spectral width of 10,000 Hz, 8192 complex points, and 128 acquisitions. All spectra were apodized using an exponential multiplication with a line broadening of 3 Hz prior to Fourier transformation.

(b) *NOESY spectra of thrombin aptamer in ^2H_2O*

A NOESY spectrum of the thrombin-binding aptamer in ^2H_2O is shown in Figure 3. The spectrum shows one set of resonances for all protons, indicating that a single conformation is present under these conditions. The well-resolved NOESY crosspeaks allowed for almost complete resonance assignments and accurate quantification of a large number of NOESY crosspeaks.

Although the NOESY spectra in ^2H_2O and H_2O are indicative of a single conformation for the thrombin aptamer, the stoichiometry of the structure could not be known *a priori*; i.e. the spectra do not allow a distinction between a monomer or a symmetric dimer (or tetramer) structure. In our preliminary report on the thrombin aptamer structure (Macaya *et al.*, 1993), we presented evidence that the thrombin aptamer structure is unimolecular. Two thrombin aptamer derivatives (**I2** and **I8**) were mixed in an equimolar ratio. The NOESY spectrum of this mixture was compared with a superposition of the NOESY spectra of **I2** and **I8** alone. The spectrum of the mixture was essentially identical to the superimposed spectra of the individual derivatives, indicating that the structure must be unimolecular, since **I2**/**I8** heterodimers would be expected to have different chemical shifts of at least some crosspeaks. In addition, both the optically (A_{276}) and the differential scanning calorimetry determined melting temperature values were concentration independent (J. A. Roe, unpublished results). Thus, a thrombin aptamer monomer was used in all of the structure calculations.

(c) *Sequence-specific assignments of non-exchangeable resonances using aptamer derivatives*

Since the aptamer structure contains both *syn* bases and loops, sequential connectivities from base

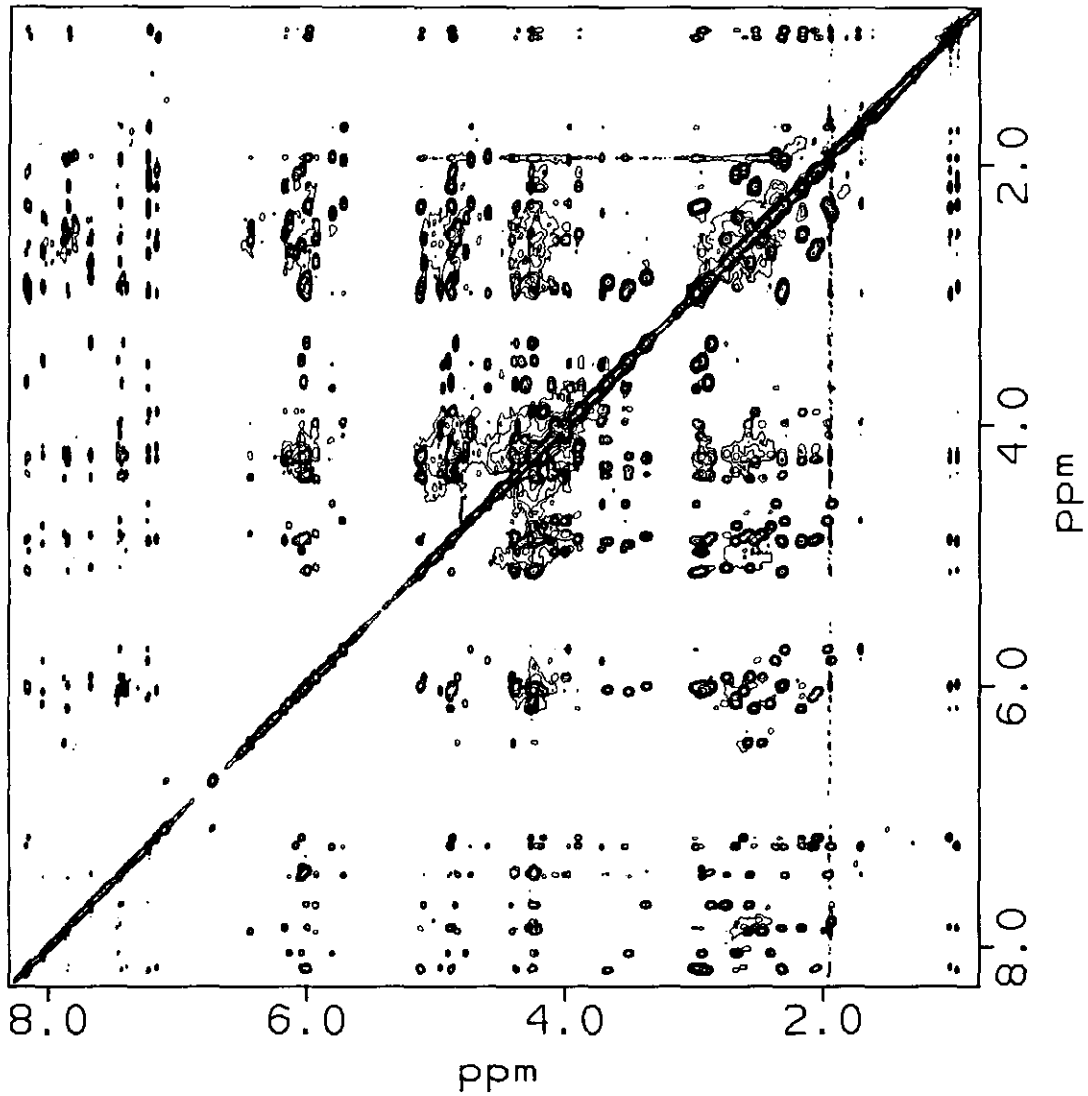

Figure 3. NOESY spectrum of the thrombin aptamer in ^2H_2O at 20°C and t_m = 250 ms. Sample was 2·3 mM strand, 110 mM KCl, in 400 μl ^2H_2O (pH 6·1). Spectrum was acquired with a spectral width of 5000 Hz in both dimensions, 2048 complex points in t_2, 32 scans per t_1 value, 300 t_1 increments, and a recycle delay of 1.8 s. The spectrum was apodized with a skewed squared sine bell (65°, 0·75, 700) in t_2 and (75°, 0·8, 300) in t_1. The final matrix size was 2048 × 2048 real points.

(H8,H6) to sugar (H1′, 2′, 2″, 3′) can only be found for short stretches in the sequences (Macaya *et al.*, 1993). For example, there are two isolated stretches of sequential connectivities for the sequence G-G-T-T (see Fig. 3 of Macaya *et al.*, 1993), but it was not possible to determine which was G1-G2-T3-T4 *versus* G11-G12-T13-T14. In the **I2** derivative, the I2H8 is identified by its low-field chemical shift. Thus, from the sequential connectivities, G1-I2-T3-T4 could be distinguished from G11-G12-T13-T14. Since the thrombin aptamer and **I2** have similar NOESY crosspeak patterns, this also allowed identification of these residues in the thrombin aptamer. Similarly, **I8** was used to assign T7, G8, and T9, and **I6** was used to differentiate between G5-G6 and G14-G15. Finally, NOE crosspeaks were observed from G2H2′,H2″ to both T3 and T4 and from G12H2′,H2″ to both T13 and T14. The **A3** derivative was used to distinguish between T3 and T4. This also resolved any ambiguity about the assignment of T3 and T4 *versus* T13 and T14. In the NOESY spectrum of **A3**, all of the NOE crosspeaks to the T3Me are missing, while the rest of the TMe crosspeaks appear at close to the corresponding chemical shifts in the thrombin aptamer. The similarity of the NOESY spectra of **A3** and the thrombin aptamer indicates that both molecules fold into the same overall structure.

(d) *Sequence specific assignments of non-exchangeable resonances from ^{31}P-^1H COSY*

A ^{31}P-^1H COSY spectrum is shown in Figure 4. Since the spin systems in each sugar had been identified, this spectrum provides direct model-independent sequential assignments of the thrombin

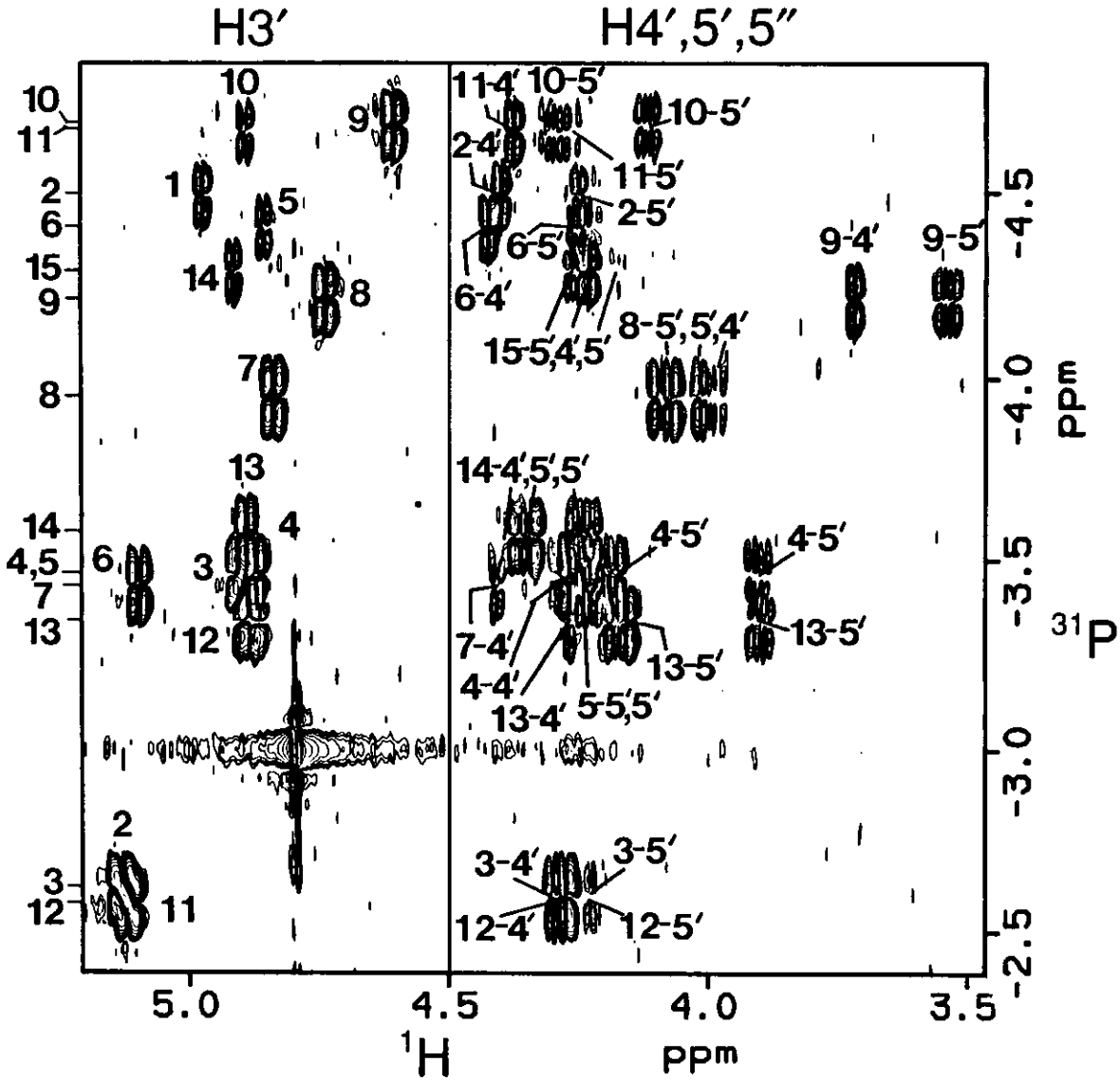

Figure 4. ^{1}H-^{31}P COSY spectrum of the thrombin aptamer at 20°C. Sample was the same as in Fig. 3. Phosphorus resonances are identified by nucleotide number on the left side of the spectrum and crosspeaks between P and H-3′,H4′ and H5′,H5″ resonances are identified on the spectrum. The H5′,H5″ resonances are not stereospecifically assigned. The spectrum was acquired with a spectral width of 1500 and 860 Hz for the ^{1}H and ^{31}P dimensions, respectively, 1024 complex points in t_2, 60 t_1 values, and 112 scans per t_1 increment. The spectrum was apodized with a skewed squared sine bell (50°, 0·7) in t_2 and (60°, 0·9) in t_1. The final matrix size was 1K × 1K real points.

aptamer. Assignments of the sugar resonances are indicated on the spectrum. Each phosphorus resonance has crosspeaks to H4′,H5′,H5″ of the 3′ nucleotide and to the H3′ of the 5′ nucleotide. Thus, sequential connectivities can be traced along the DNA strand without any structural assumptions (Pardi *et al.*, 1983; Smith & Feigon, 1993). In contrast to the direct methods used here, Wang *et al.* (1993) assigned the non-exchangeable resonances of the thrombin aptamer by model building and comparison with NOE data. The chemical shifts of the non-exchangeable proton and phosphorus resonances are given in Table 1.

(e) *Assignments of the exchangeable proton resonances*

The exchangeable proton resonances were assigned from NOE crosspeaks with non-exchangeable proton resonances and by comparison of the thrombin aptamer spectra with those obtained on the inosine-, uracil- and adenine-substituted derivatives. In a guanine quartet, NOEs are expected between the imino and amino protons of one guanine and the H8 of a neighboring guanine (Smith & Feigon, 1992, 1993; and Fig. 1). Knowledge of the H8 assignments is insufficient to assign the imino resonances to specific bases in a quartet, however, unless the overall topology of the structure is already known. This is because the relative location of the specific bases depends on the structure; e.g. the neighbors in the G-quartets would be different if the T-G-T residues looped diagonally across a quartet *versus* along an edge. The guanine neighbors in the quartets were determined using the inosine-substituted derivatives **I2**, **I6**, and **I11**. For

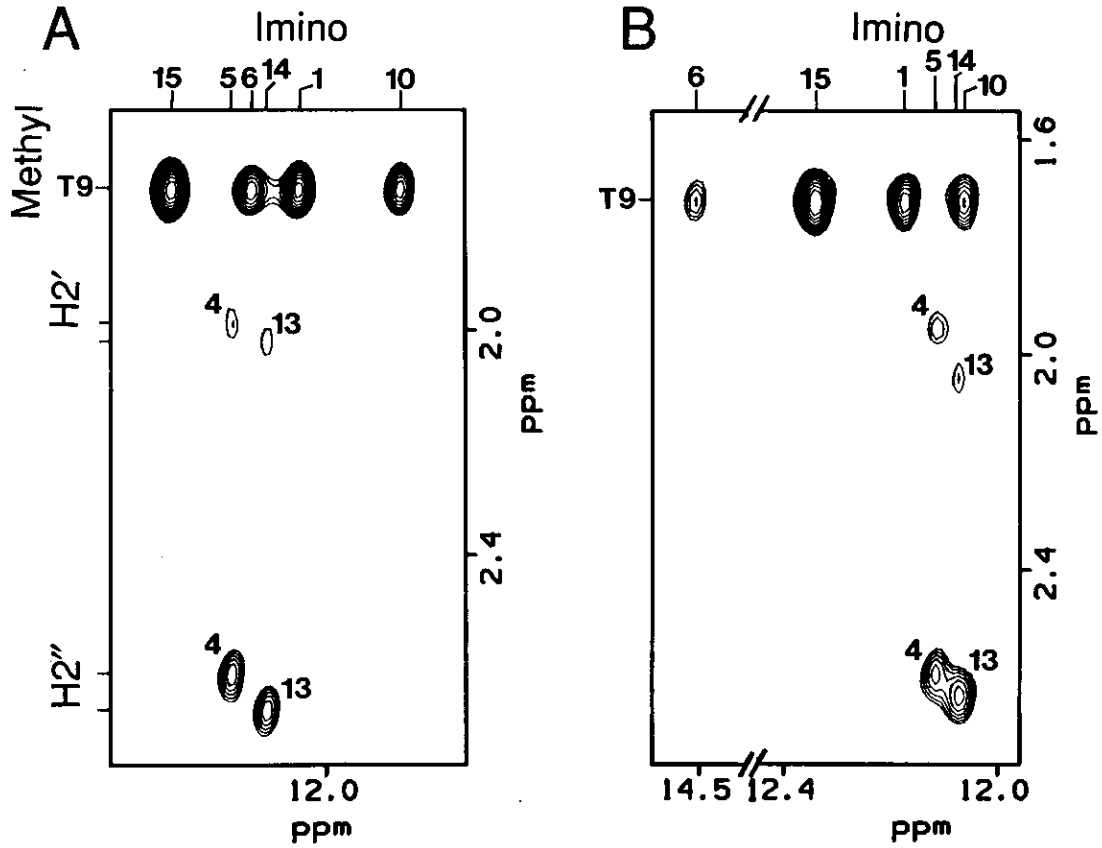

Figure 5. Portion of NOESY spectra in 90%H_2O/10%^2H_2O at 1°C of (A) thrombin aptamer and (B) **I6**. Region shown contains crosspeaks between hydrogen-bonded G and I imino proton, and T methyl and H2′, H2″ resonances. Assignments are indicated on the spectra. Sample (A) is the same as that in Fig. 3 (except for 90%H_2O/10%^2H_2O); sample (B) is under the same conditions but 2.0 mM strand. NOESY spectra were acquired with a $1\bar{1}$ spin echo read pulse (Sklenář & Bax, 1987). Spectral parameters were: (A) excitation maximum at 11·5 p.p.m., spectral width 8333 Hz in both dimensions, $t_m = 160$ ms, 1024 complex points in t_2, 300 t_1 values, 64 scans per t_1 increment, and recycle delay of 1·8 s. Spectrum (B) was acquired with the same parameters except excitation maximum at 13·3 p.p.m., spectral width of 10638 Hz, and $t_m = 210$ ms. Spectrum (A) was apodized with a skewed squared sine bell (75°, 0·8) in t_2 and (80°, 0·8) in t_1; spectrum (B) was the same but 70°, 0·7 in t_2 and 75°, 0·9 in t_1. The final matrix size was 2048 × 2048 real points.

Table 1

^1H and ^{31}P resonance assignments for the thrombin aptamer

Base	H8,H6	Methyl	H1′	H2′	H2″	H3′	H4′	H5′,5″†	Imino‡	Amino‡	^{31}P§
G1	7·39	–	6·04	2·95	2·95	4·97	4·37		12·05	9·33, 6·26	–
G2	8·16	–	5·99	3·00	2·33	5·12	4·39	4·24	12·10	10·04, 6·65	−4·49
T3	7·84	1·95	6·17	2·17	2·53	4·88	4·27	3·90	11·30	–	−2·62
T4	7·16	1·03	6·04	2·04	2·62	4·87	4·26	4·16, 3·89	11·35	–	−3·46
G5	7·42	–	6·00	3·37	2·87	4·85	4·38	4·26, 4·26	12·17	9·43, 6·75	−3·46
G6	7·68	–	5·93	2·75	2·57	5·09	4·42	4·22	12·13	10·10, 6·66	−4·40
T7	7·88	1·95	6·43	2·48	2·59	4·83	4·40		11·20	–	−3·43
G8	7·45	–	5·72	1·97	2·30	4·73	3·97	4·01, 4·08	10·68	5·80, 5·80	−3·94
T9	7·23	1·71	5·80	1·94	2·37	4·60	3·71	3·53, 2·99	10·33	–	−4·21
G10	7·43	–	6·03	3·67	2·89	4·89	4·28	4·31, 4·12	11·86		−4·68
G11	8·18	–	5·98	2·94	2·31	5·11	4·37	4·26	12·02	10·08, 6·65	−4·66
T12	7·85	1·95	6·16	2·17	2·53	4·89	4·27	3·90	11·30	–	−2·57
T13	7·22	0·97	6·08	2·08	2·68	4·89	4·27	4·18, 3·90	11·30	–	−3·34
G14	7·45	–	6·04	3·51	2·94	4·91	4·40	4·35, 4·24	12·10	9·71, 6·82	−3·57
G15	8·05	–	6·13	2·67	2·41	4·77	4·22	4·17	12·27	9·43, 6·70	−4·28

All non-exchangeable proton resonances were assigned from NOESY spectrum in ^2H_2O at 20°C, 110 mM KCl (pH 6·1).

† The H5′,5″ protons are not stereospecifically assigned.

‡ Imino and amino assignments obtained from NOESY spectrum in H_2O at 1°C. The lower-field amino resonance corresponds to the hydrogen-bonded amino proton.

§ Referenced to TMP at 20°C.

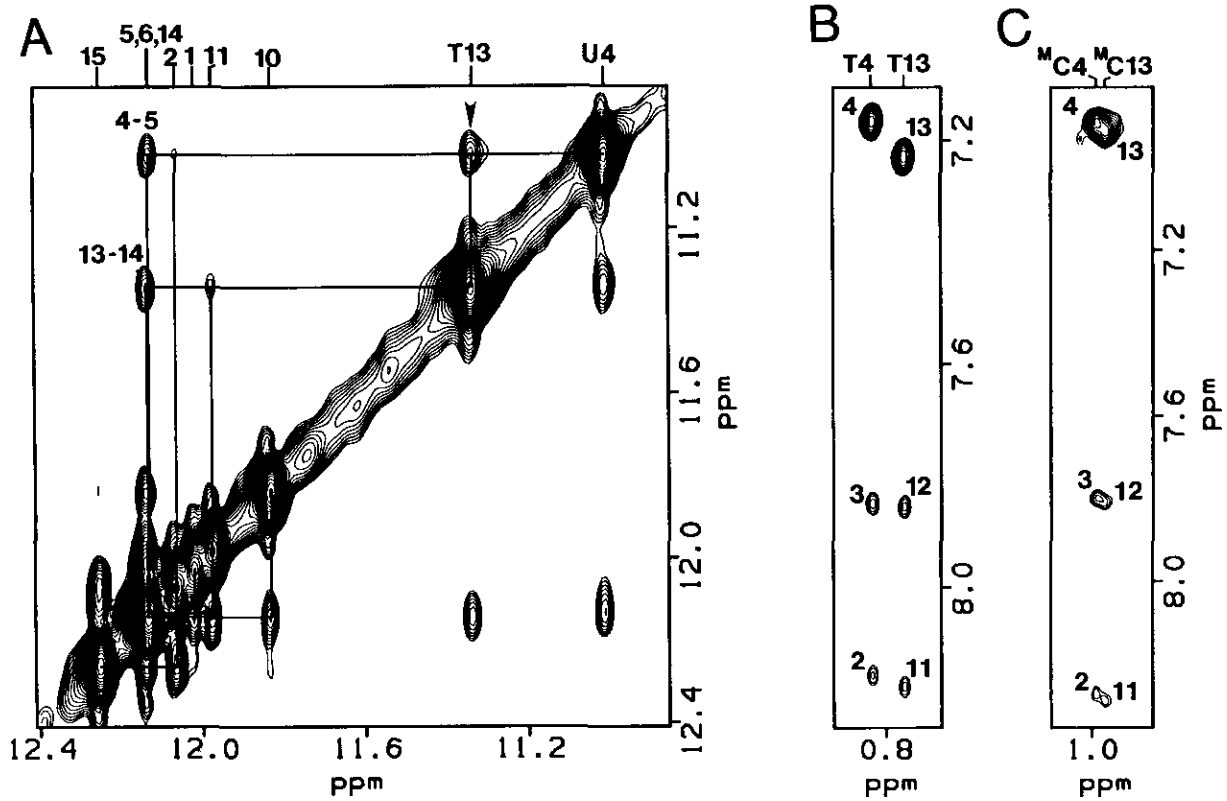

Figure 6. (A) Portion of NOESY in H_2O of **U4** at 1 °C and $t_m = 160$ ms containing the imino-imino crosspeaks. The imino assignments are indicated at the top. Imino-imino crosspeaks T13-G14 and U4-G5 are labeled; the U4-T13 crosspeak is identified by the arrowhead. The sample was 2·2 mM DNA strand, 100 mM KCl, (pH 6·0 in 400 μl 90%H_2O/10%^2H_2O. Spectrum was acquired and processed as in Fig. 5A. Due to the $1\bar{1}$ echo read pulse, the intensities of the crosspeaks are not symmetric across the diagonal. (B) Portion of NOESY spectrum in ^2H_2O of thrombin aptamer showing the region containing crosspeaks between T4 and T13 methyl and aromatic proton resonances. Sample is the same as in Fig. 3. (C) Portion of NOESY spectrum in ^2H_2O of **C4C13** showing the same crosspeak region as in (B). Sample conditions are the same as in (B), but 2·0 mM strand. Spectra were acquired and processed as in Fig. 3.

example, in NOESY spectra of the **I2** derivative, there is an NOE between the unique I2H2 and G14H8; this indicates that these two bases are neighbors in a quartet, and by analogy G2 and G14 are base-paired (Macaya *et al.*, 1993, Fig. 4). Similarly, the **I6** derivative has an NOE between I6H2 and G10H8, indicating that I/G6 and G10 are base-paired, and spectra of **I11** indicate that I/G11 and G5 are base-paired. The only topology that is consistent with these results is that shown in Figure 1, with the two G-quartets G1·G6·G10·G15· and G2·G5·G11·G14·.

In the NOESY spectrum of the thrombin aptamer in H_2O, there are NOE crosspeaks between the T9 methyl and four different hydrogen-bonded G-imino resonances (Fig. 5A). This is only possible for four guanine residues in the same quartet. The **I6** inosine imino also shows a crosspeak to the T9 methyl (Fig. 5B). This indicates by analogy that G6 is in this quartet, and from knowledge of the quartet structure obtained as discussed above so are G1, G10 and G15.

T4 and T13 imino protons have NOEs with G5 and G14 imino protons, respectively. The assignments of these as sequential imino-imino NOEs was confirmed with the **U4** derivative (Fig. 6). Although weak imino-imino NOE crosspeaks are observed between T3 and T4 and between T12 and T13 in the NOESY spectrum of **U4**, these crosspeaks could not be resolved in the thrombin aptamer due to spectral overlap, so the assignment of T3 and T12 imino resonances is ambiguous. The non-hydrogen-bonded G8 imino was identified by its strong intrabase imino-amino NOE crosspeak. The T9 imino was distinguished from the T7 imino resonance on the basis of NOE crosspeaks between T9 imino and G15H1′ and H2″ resonances. Chemical shifts of the exchangeable proton resonances are given in Table 1.

(f) *NOESY spectra of U4 and C4C13 derivative: evidence for T4·T13 base-pair*

In the imino region of the NOESY spectrum of the thrombin aptamer, NOE crosspeaks are observed between G imino and T4 and T13 imino protons. Since non-hydrogen-bonded imino protons would usually exchange too fast to have NOEs, this raised the possibility that there might be a T4·T13 base-pair. Such a base-pair could form by hydrogen bonding between either the O2 or the O4 carbonyls to the N3 hydrogens of the two thymines (Fig. 7). In

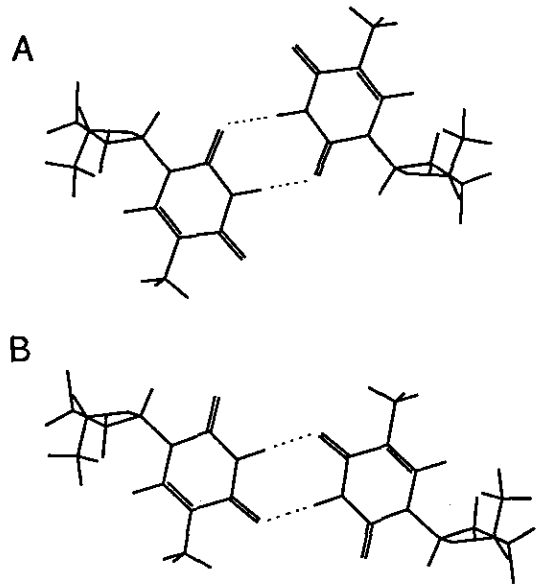

Figure 7. Schematic drawing of (A) O2 and (B) O4 T-T base-pairs. The geometry of base-pairing was chosen so that the N–H···O configuration is linear and the C=O···H angle is near 120°.

either case, an NOE would be expected between the two thymine imino protons. In the thrombin aptamer, the chemical shifts of T4 and T13 imino proton resonances are too close to resolve such an NOE. In order to resolve this, the **U4** derivative was synthesized. In the imino-imino region of the NOESY spectrum of **U4** shown in Figure 6A, imino-imino NOE crosspeaks are observed between T13 and U4, as well as between U4-G5, U4-G2 (weak), T13-G14, and T13-G11 (weak). Crosspeaks are also observed between the imino proton resonances of G5 and G14 and the H2″ (and H2′) resonances of T4 and T13, respectively (Figure 5), and between T4Me-G2H8 and T13Me-G11H8 (Fig. 6B). These crosspeaks indicate that T13 and T4 (U4) are stacked on the G2·G5·G11·G14· quartet and are probably base-paired. It is not possible from the data to distinguish between the O2 and the O4 hydrogen bonding.

In order to try to distinguish between O2 and O4 hydrogen bonding for the T·T base-pair, the **C4C13** derivative, in which T4 and T13 are replaced by a m^5C was made and studied by NMR. The m^5C was chosen as a substitute for T, since the O4 is now replaced by an amino group. This would preclude O4 base-pairing, but would allow O2 pairing at low pH. The NMR spectra of **C4C13** indicate that the stacking of C4 and C13 is virtually identical to that of T4 and T13; i.e. both the crosspeak patterns and intensities are the same for the two molecules (Fig. 6B, C). However, we were unable to observe protonated C imino resonances unambiguously. Similar results were obtained with **U4C13**. Thus, these results favor but do not prove that the T4 and T13 are hydrogen bonded *via* O2.

(g) *Initial sugar conformations and backbone ε, β, and γ dihedral bond angle restraints*

The calculated sugar proton coupling constants and %S-type sugar conformation are given in Table 2. All of the sugars are predominantly S-type.

For all sugars in the thrombin aptamer, all $J_{P,5'}$ and $J_{P,5''}$ were < 8 Hz. Using the Karplus relation (Lankhorst *et al.*, 1984), the dihedral angles (P-O5′-C5′-H5′) and (P-O5′-C5′-H5″) are each restricted to the two ranges $70 \pm 30°$ and $-70 \pm 30°$. However,

Table 2

Coupling constants and sugar conformations of deoxyribose rings in thrombin aptamer derived from H1′-H2′ and H1′-H2″ crosspeaks using CHEOPS

	$J_{1'2'}$	$J_{1'2''}$	$J_{2'3'}$	$J_{2''3'}$	Corr‡	%S§	P_N‖	P_S‖
G1†	–	–	–	–	–	–	–	–
G2	11·70	6·30	7·10	0·70	0·9632	100		121
T3†	–	–	–	–	–	–	–	–
T4	10·30	5·90	3·00	3·40	0·9502	88¶		171
G5	9·00	4·40	4·40	0·30	0·9441	100¶		167
G6	12·30	3·30	3·20	1·70	0·9599	100¶		159
T7	9·30	5·00	5·00	1·00	0·9494	92		154
G8	8·00	4·70	4·80	4·10	0·9229	70	−15	159
T9	9·40	5·30	4·10	1·70	0·9308	92		172
G10	9·80	4·30	1·80	0·60	0·9697	100		173
G11	10·40	5·60	8·20	0·10	0·9685	100		115
T12†	–	–	–	–	–	–	–	–
T13	10·90	7·00	5·60	1·20	0·9339	69	107	180
G14	10·00	5·20	1·10	0·10	0·9755	73	69	171
G15	10·90	6·50	4·30	1·20	0·9198	100		176

† G_1H2' and H2″ have same chemical shift. T3 and T12 resonances overlap.

‡ Correlation coefficient.

§ S-type sugar conformation calculated as described.

‖ Pseudorotation angle calculated for S-type and N-type sugars. The P_N value was not considered sufficiently determined if %S was ≥ 75.

¶ RMS difference between CHEOPS coupling constants and PSEUROT coupling constants >1, perhaps due to the lesser accuracy in the determination of $J_{2'3'}$ and $J_{2''3'}$ couplings.

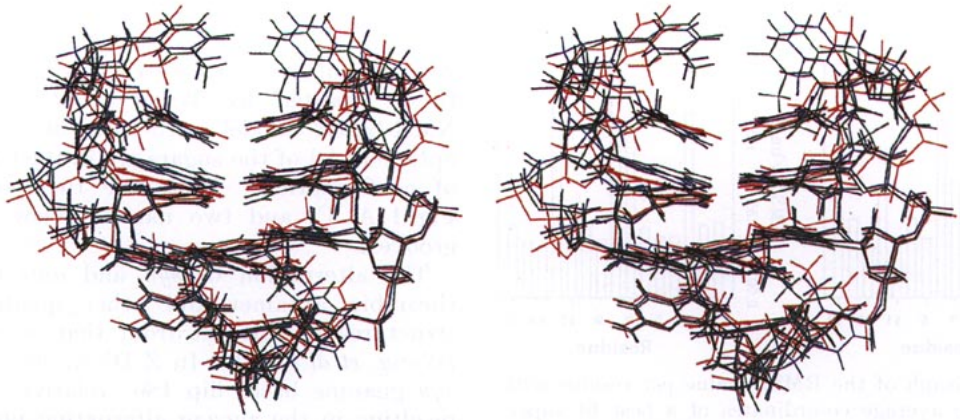

Figure 8. The lowest energy conformation for O0 (blue), O2 (red) and O4 (green) calculations of thrombin aptamer structure. Orientation in the stereoview is: wide groove in front, 2 T-T connecting loops at top, with T4 and T13 seen edge on.

when the additional constraint that the dihedral angles (P-O5′-C5′-H5′) and (P-O5′-C5′-H5″) must differ by about 120° is considered, the range of β is restricted to $180 \pm 30°$, even without stereospecific assignments.

For residues 3, 4, 6, 7, 8, 9, 12 and 13, $J_{P,H3'}$ was between 8 and 11 Hz. The corresponding ε angles were restrained to $-120 \pm 45°$. According to the Karplus relation, ε angles near 0° and 120° would likewise result in $J_{P,H3'}$ values around 10 Hz. However, these values are not sterically feasible and do not occur in crystal structures of 45 di- and trinucleotides (Sundaralingam & Haromy, 1989). For residues 1, 5, 10 and 14, which have $J_{P,H3'} < 6$ Hz, the ε restraints used were $-120 \pm 110°$. Coupling constants for G2 and G11 could not be reliably determined due to overlap of the involved crosspeaks, so no ε restraints were used for these nucleotides.

The $\Sigma J_{H4'}$ used in estimating γ were all <10 Hz. This sum depends on both γ and δ. A plot of $\Sigma J_{H4'}$ *versus* γ and the pseudorotation phase angle P (on which δ depends directly) (fig. 3 of Kim *et al.*, 1992) shows that for all values of $\Sigma J_{H4'} < 10$ Hz, γ must fall into one of the two ranges $60 \pm 40°$ and $240 \pm 40°$. For bases in the *anti* conformation, the latter range can be excluded on the basis of the absence of strong H6/8-H5′,H5″ NOE crosspeaks (Kim *et al.*, 1992). However, for nucleotides in the *syn* conformation, these crosspeaks could be absent for all possible γ values, so no distinction can be found between the two ranges. Therefore, no dihedral restraints for γ were applied to the four G nucleotides in the *syn* conformation.

(h) *Refinement of thrombin aptamer structure*

Three different sets of 20 structure calculations were performed, which differed only in the constraints between T4 and T13: (1) no explicit hydrogen bonds (O0), (2) explicit hydrogen bonds between O2 and N3 (O2), and (3) explicit hydrogen bonds between O4 and N3 (O4). The average RMSD values after NOE relaxation matrix refinement for the six best (lowest overall energy) structures calculated for the heavy atoms of all nucleotides are 1·56 Å (O0), 1·00 Å (O2) and 1·15 Å (O4).

Three structures corresponding to the best structure (lowest energy) from each set are shown superimposed in Figure 8. The three sets of calculations do not allow an unambiguous determination of the relative geometry of T4 and T13. Structures from all three sets of calculation are sterically feasible and satisfy the NOE constraints. The variation in total energy values between the sets is within the variation within each set.

The average R-factor for the six best O2 structures, calculated with an exponent of 1/6 and otherwise analogous to the crystallographic residual (Brünger, 1992), was 0·113 before and 0·061 after the direct NOE relaxation matrix refinement. Although the direct NOE relaxation matrix refinement significantly improved the agreement between measured and calculated NOE intensities, it did not lead to very significant conformational changes. For the ensemble of the six best O2 structures, the average RMSD decreased from 1·25 to 1·00 Å (calculated for all heavy atoms).

In the refined structures, all of the nucleotides except T3, T7, T8 and T12 are well defined. This is clearly illustrated in the RMSD *versus* residue plots for the six best structures from the O2 set in Figure 9. The positions of T3, T7, G8 and T12 are relatively poorly constrained by the NMR data; these nucleotides have an average of 12 distance constraints *versus* the average of 25·7 for the rest of the nucleotides. The average RMSD values calculated excluding T3, T7, G8 and T12 (superposition without the excluded residues), are 1·12 Å (O0), 0·62 Å (O2) and 0·75 (O4), significantly lower than those calculated including all nucleotide heavy atoms.

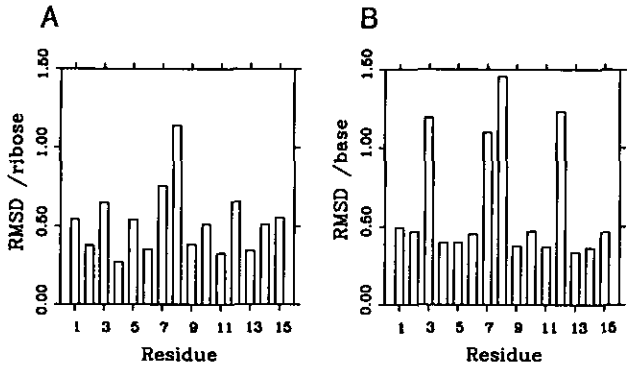

Figure 9. Graph of the RMSD value per residue with respect to the average co-ordinates of a best fit superposition using the non-hydrogen backbone atoms of the ensemble of the 6 lowest energy structures of the O2 calculations for the thrombin aptamer. Values are plotted for the non-hydrogen atoms of the (A) deoxyriboses and (B) bases.

(i) *Backbone dihedral angles in refined structures*

Figure 10 illustrates the backbone dihedral angles for the six best O2 structures. Making appropriately conservative use of the available information on backbone dihedral angles leads to improved local definition of the backbone. Since no experimental information is available for restraining α and ζ, their values are widely scattered. All values of β fall in the *trans* range usually observed in crystal structures. The γ values for all of the *anti* nucleotides fall in a narrow range around 60° due to restraints derived from H4′–H5′,H5″ coupling constants and relative NOE intensities (Kim *et al.*, 1992). The γ values for the *syn* residues could not be constrained by the relative NOE intensities, and therefore fall into three allowed ranges separated by 120° due to steric hindrance for rotation around the C5′–C4′ bond. The fairly narrow range of δ values for each nucleotide reflects the ν_1 and ν_2 constraints on the deoxyribose ring conformation from ^1H–^1H coupling constants and NOEs. The ε values fall into the allowed range found in the crystal structures and also by Hilbers and co-workers (Mooren *et al.*, 1993). Finally, the χ values reflect the *syn* conformation of G1, G5, G10 and G14 and *anti* conformation of the rest of the nucleotides.

4. Discussion

(a) *Structural features of the thrombin aptamer core*

Superpositions of the six best structures from the O2 set are shown in Figure 11 along with one view of a single structure for clarity. The thrombin aptamer forms a right-handed helix with two stacked guanine quartets and a regular phosphodiester backbone. The guanine quartets are linked by two T-T loops (T2-T3 and T12-T13) at one end and a T-G-T loop at the other end. Each guanine run is 5′-*syn*-*anti*-3′, while all of the thymines and G8 in the loop are *anti*. The guanines in the two guanine quartets alternate glycosidic torsion angles (and orientation) around each quartet: $G1_{syn}$·$G6_{anti}$·$G10_{syn}$·$G15_{anti}$ and $G2_{anti}$·$G5_{syn}$·$G11_{anti}$·$G14_{syn}$. These general features of the structure are the same as those reported by Wang *et al.* (1993) for an NMR-based model structure of the thrombin aptamer. All of the sugars are in the C2′-*endo* range of conformations. There are two wide (view in Fig. 11A, C) and two narrow (view in Fig. 11B) grooves.

The alternation of *syn* and *anti* bases in the thrombin aptamer and other quadruplex DNA structures is different from that seen in Z-DNA (Wang, *et al.*, 1979). In *Z*-DNA, the sugars of the *syn* guanine bases flip 180° relative to the bases, resulting in the zig-zag alternating phosphodiester backbone, and the *syn* guanines have C3′-*endo* sugar puckers (Rich *et al.*, 1984). In the thrombin aptamer, the *syn* guanine bases flip relative to the sugars, so the phosphodiester backbone remains regular as in *B*-DNA. The narrow range of ^{31}P chemical shifts of the guanine phosphates (Fig. 4) is also consistent with a *B*-DNA-like phosphodiester backbone.

The guanine quartets show some deviation from planarity in a nonrandom, systematic way among the ensemble of the six best conformations (Fig. 8 and 11A, B). The top G-quartet (G2·G5·G11·G14·) appears more closely planar than the bottom G-quartet. The guanine bases in the quartets are the best defined part of the structure, while more variability is seen in the sugars and especially the phosphodiester backbone. This variability can be partially attributed to the fact that there are few constraints on the phosphodiester backbone, as well as to the inherent dynamics of the ribose conformations, which is not accounted for in the structure calculations. This inherent dynamics is evidenced by the fact that most of the observed *J*-coupling constants cannot be fully explained by a single static ring conformation.

The stacking of the guanine quartets is illustrated in Figure 11D. This view also shows the two wide grooves and the two narrow grooves of the quadruple helix. The average helical twist is small but the numerical value depends on how one defines it; the problem in defining it arises from the fact that the bases alternate orientation from one quartet to the next. Average twist values are around 20° and 35°, as measured by the positions of the C1′ atoms or the geometric centers of the bases, respectively. The average distance between the two guanine quartets is 3·8 Å.

(b) *Structural features of the loops*

The two T-T loops at one end of the quadruplex each span a narrow groove, while the T-G-T loop at the other end of the quadruplex spans a wide groove. T9 is stacked rather flatly on the neighboring G-quartet and shows as little conformational variability as the tetrads (Fig. 9). G8, although less well defined, is also stacked on the neighboring G-quartet. T7 is generally positioned sideways into

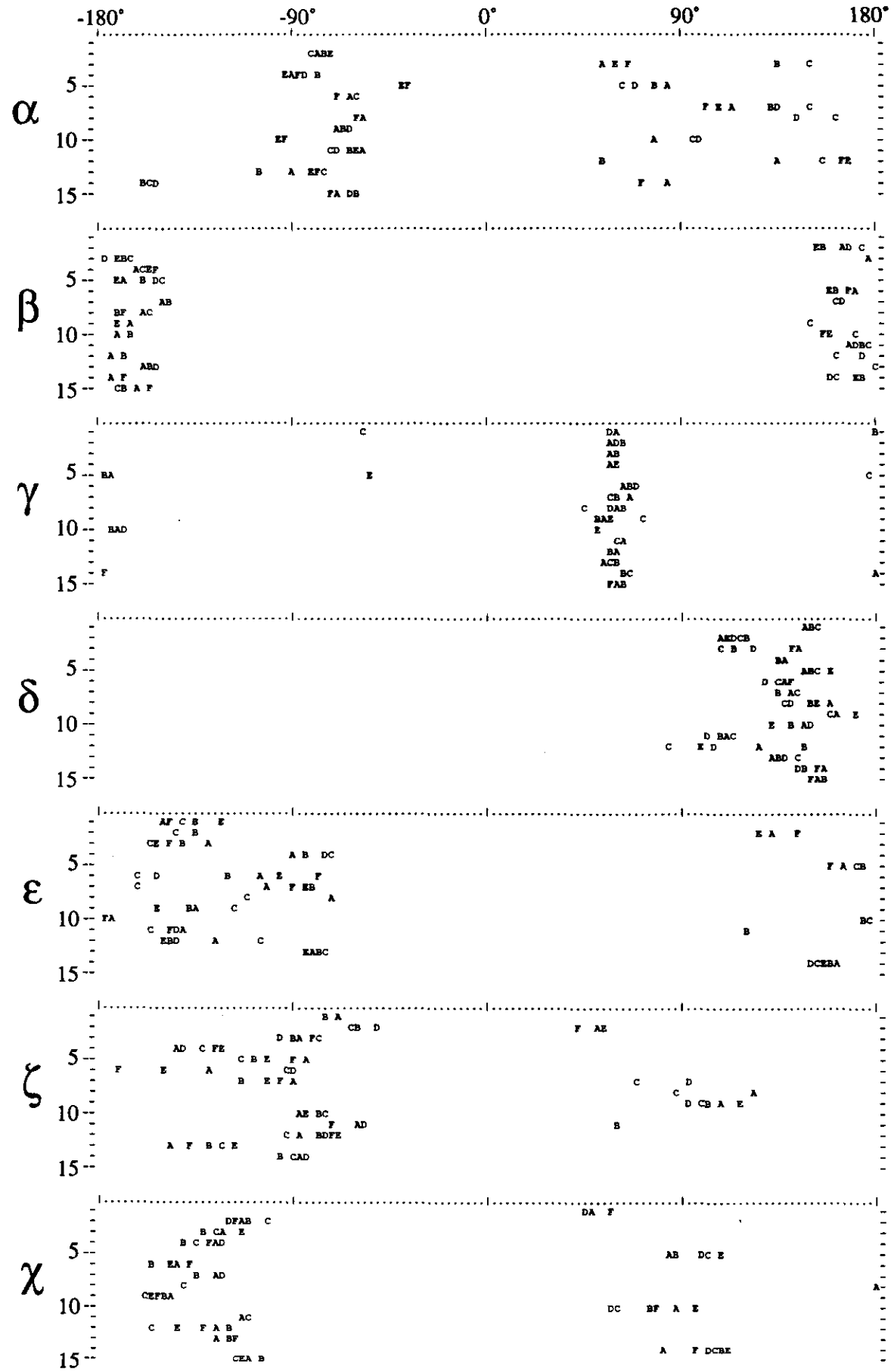

Figure 10. Graphical representation of all values for α, β, γ, δ, ε, ξ and χ for the same structures as in Fig. 11. The angles are plotted in 120 intervals of 3° *versus* the sequence positions and denoted with letters A through F designating the 6 conformations in ascending order of their total energies. For clarity, in cases where 2 or more values overlap, only 1 letter is plotted.

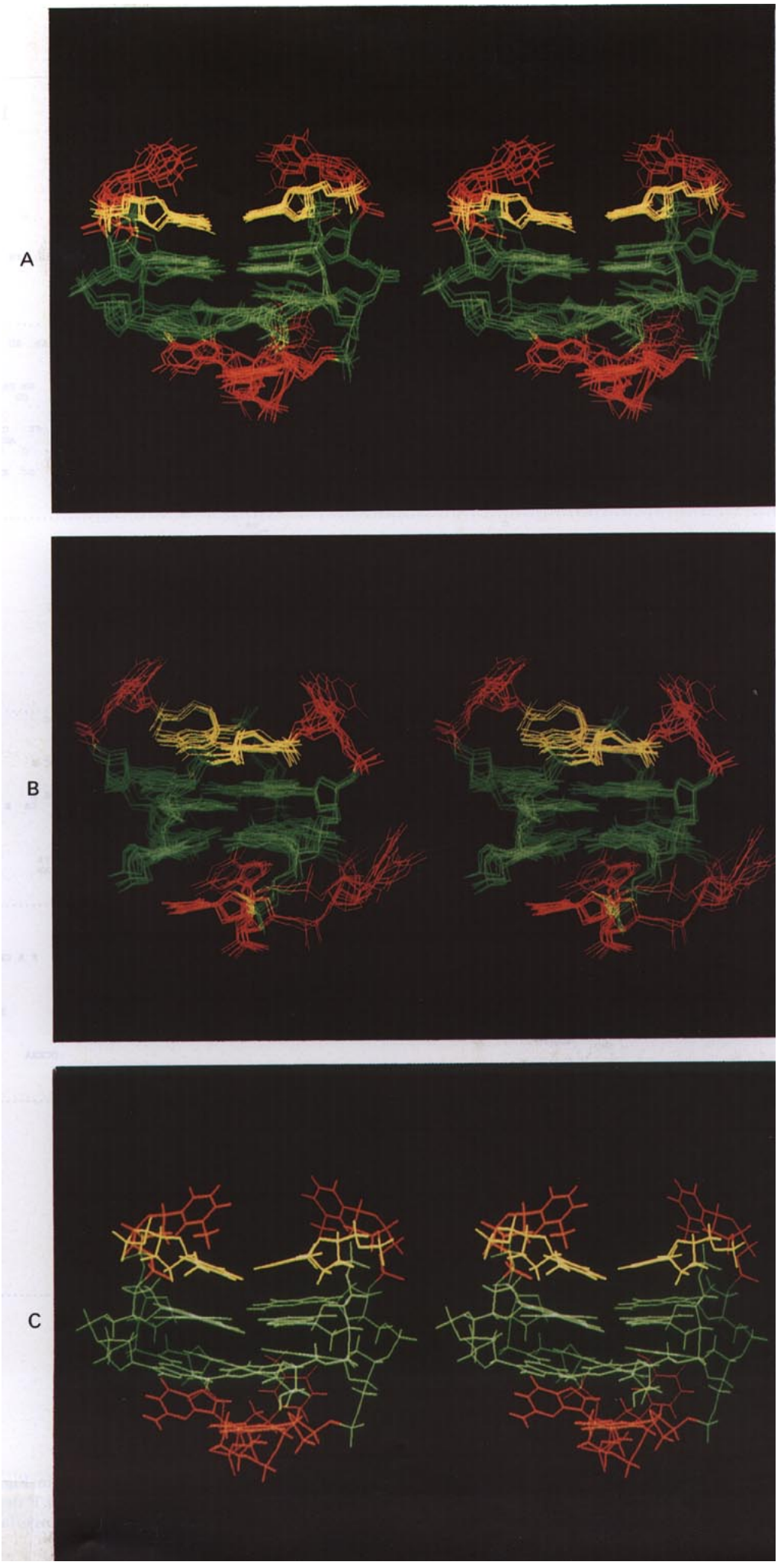

Fig. 11.

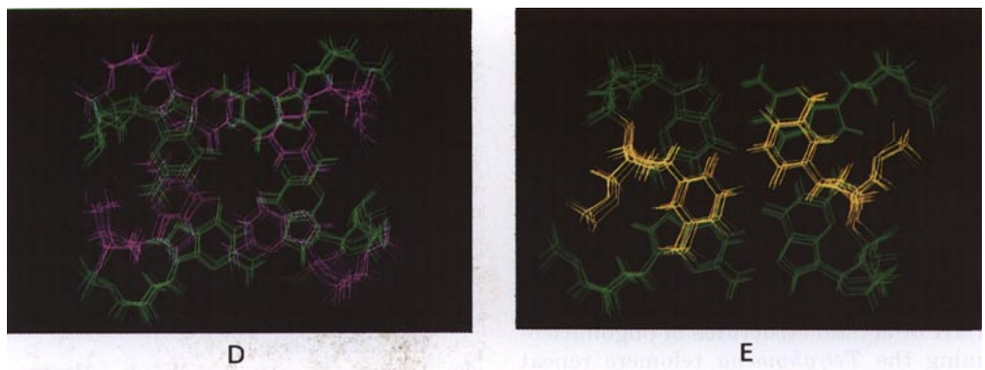

Figure 11. Stereoviews of the 6 best O2 structures of the thrombin aptamer: (A) View into the wide groove. (B) View into the narrow groove. (C) View into wide groove of the best O2 structure. The 2 G-quartets are green, the T4·T13 base-pair is yellow, and the other nucleotides are red. (D) Stacked G-quartets: bottom quartet, clockwise from lower left, G1·G6·G10·G15 (purple); top quartet, clockwise from lower left, G2·G5·G11·G14 (green). (E) Stacking of T4·T13 (yellow) on G2·G5·G11·G14 (green) quartet.

the wide groove spanned by this loop. This can be most clearly seen in Figure 11B.

In the T3-T4 and T12-T13 loops, both T4 and T13 are well defined and are stacked on their neighboring G-quartet. In contrast, T3 and T12 show a high degree of variability in the ensemble of well-converged solution structures (Figs 9 and 11A, B). In general they are positioned so that thymine rings are tilted away from the neighboring T bases, with the methyl group pointing down toward the wide groove. The position of the aromatic rings of T3 and T12 relative to the methyl groups of T4 and T13, respectively, is consistent with unusually high-field chemical shifts for the T4 and T13 methyl resonances (Table 1), which would be expected due to ring current effects.

The stacking of T4 and T13 on the neighboring G-quartet positions these nucleotides so that they could easily be within hydrogen-bond distance to form a T·T base-pair. Although the expected imino-imino NOE for a T·T base-pair cannot be unambiguously resolved in the thrombin aptamer, due to the close chemical shifts of T4 and T13 imino resonances, there is a relatively strong imino-imino NOE between the resolved U4 and T13 in **U4** (Fig. 6A). Taken together, this is strong evidence for formation of a T·T base-pair in the thrombin aptamer. In addition, in the consensus sequence for thrombin binding aptamers, the bases corresponding to T4 and T13 are invariant, consistent with base-pairing of T4 and T13, while T3 and T12 are variable. Replacement of T3 with A3 in the **A3** derivative does not disrupt the tertiary structure of the aptamer (Fig. 2). The NMR spectra of **C4C13** and **U4C13** support the O2 bonding scheme.

The structure calculations do not provide a clear distinction between a T·T base-pair hydrogen bonded at O4, a T·T base-pair hydrogen bonded at O2, or simple stacking. This is because a relatively small shift in the phosphodiester backbone will accommodate each of the three possibilities and there are relatively few constraints between the two Ts and the neighboring G-quartet. Furthermore, many of these constraints are from exchangeable proton resonances and thus are not tight constraints. Both the O2 and O4 structures have significantly lower average RMSD values than the O0, with the O2 structures having the lowest average RMSD. The stacking of T4 and T13 on its neighboring G-quartet is illustrated in Figure 11E for the O2 set of calculations. The T·T base-pair is to our knowledge the first structural evidence for parallel T·T mismatch. Parallel stranded U·U and C·C mismatches in rRNA have been proposed by Gutell & Woese (1990) based on phylogeny.

(c) *The quadruplex motif*

The association of guanines to form G-quartets has been known for more than 30 years (reviewed by Guschlbauer *et al.*, 1990). More recent NMR and X-ray crystallographic studies of DNA oligonucleotides that form unimolecular or bimolecular G-quadruplex structures have indicated that in these "fold-back" quadruplex structures the guanines alternate *5′-syn-anti-3′* (Kang *et al.*, 1992; Smith & Feigon, 1992, 1993; Wang *et al.*, 1991). The unexpected finding about the thrombin aptamer structure is that it forms a stable quadruplex containing only two guanine quartets. This two-quartet quadruplex is stabilized by stacking of T4 and T13 on one of the quartets as well as by the two additional hydrogen bonds in the T4·T13 base-pair and to a lesser extent by stacking of G8 and T9 on the other G-quartet.

The thrombin aptamer structure can be compared with the unimolecular and bimolecular quadruplexes formed by oligonucleotides containing the *Oxytricha* telomere repeat sequence d(TTTTGGGG), (Zakian, 1989). A unimolecular quadruplex is formed by $d(G_4T_4G_4T_4G_4T_4G_4)$ with four G-quartets and three TTTT loops; the TTTT loops

go across the edge of each of the two medium grooves at one end of the G-quartet stack, while the central TTTT loop goes across the diagonal of the other end G-quartet (Smith & Feigon, 1992, and unpublished results). A symmetric dimeric quadruplex with thymine loops at opposite ends is formed by $d(G_4T_4G_4)$. In the NMR solution structure of $d(G_4T_4G_4)$ the thymines loop across the diagonal of the G-quartets on the end (Smith & Feigon, 1992, 1993), while in the crystal structure the thymines loop across a wide groove (Kang *et al.*, 1992). No NMR or crystal structures of oligonucleotides containing the *Tetrahymena* telomere repeat d(TTGGGG) have yet been solved. The structure of the thrombin aptamer indicates that it should be possible for oligonucleotides containing two or four *Tetrahymena* telomere repeats to form a quadruplex with T-T loops connecting the G-quartets. We have obtained preliminary results on a four-quartet version of the thrombin aptamer $d(G_4TTG_4TGTG_4TTG_4)$ and find that it folds to form a unimolecular quadruplex with the same topology as the thrombin aptamer but with four G-quartets (Macaya & Feigon, unpublished results).

(d) *Relevance to thrombin binding*

Although the solution structure of the thrombin aptamer alone does not necessarily give information on the structure when bound to thrombin, it seems unlikely that there would be a large structural rearrangement of so compact a structure upon binding. The binding site on thrombin was recently localized, not surprisingly, to the anion binding exosite (Wu *et al.*, 1992). This is the site occupied by a peptide fragment, hirugen, of the thrombin-inhibiting protein hirudin. Sulfonation of a tyrosine in hirugen has been shown to increase its binding affinity by an order of magnitude (Skrzypczak-Jankun *et al.*, 1991). One might thus speculate that one of the phosphate groups of the thrombin aptamer occupies the same site as the sulfate group in the hirugen–thrombin complex. Since the T-G-T loop is variable in the consensus sequence, we also speculate that the important structural determinants for thrombin binding are in the two G-quartets and the T·T base pair in the T-T loops. It will be interesting to see if the quadruplex motif shows up in selections for aptamers that bind other proteins or ligands.

This work was supported by NIH grants R01JGM 37254-06 and R01 GM48123-01, and National Science Foundation Presidential Young Investigator Award (DMB 89-58280), with matching funds from AmGen Inc., Monsanto Co. and Sterling Winthrop Drug Inc. to J.F. and a NIH Predoctoral Biotechnology training grant (GM 08375) to R.F.M. The co-ordinates of the final structures have been deposited with the Brookhaven Data Bank and are available on request. The accession number is not available at the time of publication.

References

Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature (London)*, **355**, 564–566.

Brünger, A. T. (1992). *X-PLOR (Version 3.0) Manual*, The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT.

Davis, D. G. & Bax, A. (1985). Assignment of complex ^1H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy. *J. Amer. Chem. Soc.* **107**, 2820–2821.

de Leeuw, F. A. A. M. & Altona, C. (1983). Computer-assisted pseudorotation analysis of five-membered rings by means of proton spin-spin coupling constants: program PSEUROT. *J. Comp. Chem.* **4**, 428–437.

Ellington, A. D. & Szostak, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature (London)*, **346**, 818–822.

Feigon, J., Sklenář., V., Wang, E., Gilbert, D. E., Macaya, R. F. & Schultze, P. (1992). ^1H NMR Spectroscopy of DNA. In *Methods in Enzymology* (Lilley, D. M. J. & Dahlberg, J. E., eds), Academic Press, Inc., San Diego, CA, pp. 235–253.

Guschlbauer, W., Chantot, J.-F. & Thiele, D. (1990). Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomere DNA. *J. Biomol. Struct. Dynam.* **8**, 491–511.

Gutell, R. R. & Woese, C. R. (1990). Higher order structural elements in ribosomal RNAs *M* pseudo-knots and the use of noncanonical pairs. *Proc. Nat. Acad. Sci., U.S.A.* **87**, 663–667.

Irvine, D., Tuerk, C. & Gold, L. (1991). SELEXION–systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J. Mol. Biol.* **222**, 739–761.

Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992). Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature (London)*, **356**, 126–131.

Kim, S.-G., Lin, O.-J. & Reid, B. R. (1992). Determination of nucleic acid backbone conformation by ^1H NMR. *Biochemistry*, **31**, 3564–3574.

Kumar, A., Ernst, R. R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6.

Lankhorst, P. P., Haasnoot, C. A. G., Erkelens, C. & Altona, C. J. (1984). Carbon-13 NMR in conformational analysis of nucleic acid fragments. 2. A reparametrization of the Karplus equation for vicinal NMR coupling constants in CCOP and HCOP fragments. *J. Biomol. Struct. Dynam.* **1**, 1387–1405.

Macaya, R. F., Schultze, P. & Feigon, J. (1992). Sugar conformations in intramolecular DNA triplexes determined by coupling constants obtained by automated simulation of P.COSY crosspeaks. *J. Amer. Chem. Soc.* **114**, 781–783.

Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A. & Feigon, J. (1993). Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Nat. Acad. Sci., U.S.A.* **90**, 3745–3749.

Marion, D. & Bax, A. (1988). P.COSY, a sensitive alternative for double-quantum-filtered COSY. *J. Magn. Reson.* **80**, 528–533.

Mooren, M. M. W., Wijmenga, S. S., van der Marel, G. A.,

van Boom, J. H. & Hilbers, C. W. (1993). *The Solution Structure of the Circular Trinucleotide r(GpGpGp) Determined by NMR and Molecular Mechanics*. Ph.D. thesis, chap. 4, University of Nijmegen, The Netherlands.

Nilges, M., Habazettl, J., Brünger, A. T. & Holak, T. A. (1991). Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.* **219**, 499–510.

Pardi, A., Walker, R., Rapoport, H., Wider, G. & Wüthrich, K. (1983). Sequential assignments for the H-1 and P-31 atoms in the backbone of oligonucleotides by two-dimensional nuclear magnetic resonance. *J. Amer. Chem. Soc.* **105**, 1652–1653.

Rich, A., Nordheim, A. & Wang, A. H.-J. (1984). The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **53**, 791–846.

Sklenář, V. & Bax, A. (1987). Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson.* **74**, 469–479.

Sklenář, V., Miyashiro, H., Zon, G., Miles, H. T. & Bax, A. (1986). Assignment of the ^{31}P resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Letters*, **208**, 94–98.

Skrzypczak-Jankun, E., Carperos, V. E., Ravichandran, K. G., Tulinksy, A., Westbrook, M. & Maraganore, J. M. (1991). Structure of the hirugen and hirulog 1 complexes of α-thrombin. *J. Mol. Biol.* **221**, 1379-1393.

Smith, F. W. & Feigon, J. (1992). Quadruplex structure of telomeric DNA oligonucleotides. *Nature (London)*, **356**, 164–168.

Smith, F. W. & Feigon, J. (1993). Strand orientation in the DNA quadruplex formed from the *Oxytricha* telomere repeat oligonucleotide d(G4T4G4) in solution. *Biochemistry*, **32**, 8682-8692.

States, D. J., Haberkorn, R. A. & Ruben, D. J. (1982). A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Res.* **48**, 286–292.

Sundaralingam, M. & Haromy, T. P. (1989). Crystal structures of bases, nucleosides, and nucleotides. In *Nucleic Acids: Crystallographic and Structural Data I* (Saenger, W., ed.), Springer-Verlag, Berlin, pp. 22–254.

Szostak, J. W. (1992). In vitro genetics. *Trends Biochem. Sci.* **17**, 89–93.

Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.

van Wijk, J., Huckriede, B. D., Ippel, J. H. & Altona, C. (1992). Furanose sugar conformations in DNA from NMR coupling constants. *Methods Enzymol.* **211**, 286–306.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. & Rich, A. (1979). Molecular structure of a left-handed double-helical DNA fragment at atomic resolution. *Nature (London)*, **282**, 680–686.

Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathan, S. & Bolton, P. H. (1993). A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry*, **32**, 1899–1904.

Wang, Y., de los Santos, S., Gao, X., Greene, K., Live, D. & Patel, D. J. (1991). Multinuclear nuclear magnetic resonance studies of Na cation-stabilized complex formed by d(G-G-T-T-T-T-C-G-G) in solution; implications for G-tetrad structures. *J. Mol. Biol.* **222**, 819–832.

Wu, Q., Tsiang, M. & Sadler, J. E. (1992). Localization of the single-stranded DNA binding site in the thrombin anion-binding exosite. *J. Biol. Chem.* **34**, 24408–24412.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, NY.

Zakian, V. A. (1989). Structure and function of telomeres. *Annu. Rev. Genet.* **23**, 579–604.

Edited by P. E. Wright

(Received 20 July 1993; accepted 27 September 1993)

Note added in proof: After submission of this manuscript, a crystal structure of the thrombin aptamer bound to thrombin was published (Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E. & Tulinsky, A. (1993). *J. Biol. Chem.*, **24**, 17651–17654). The topology of the thrombin aptamer in the crystal structure appears at first glance to be the same as the solution NMR structures (this work and Wang *et al.* (1993)). Both the crystal and solution structures have two G-quartets connected by T-T and T-G-T loops. However, the strand polarity of the DNA oligonucleotide is reversed relative to the solution structure. This results in the two T-T loops spanning the wide grooves and the T-G-T loop spanning a narrow groove.